6 Ergebnissen vertrauen

6.1 Stichproben und Modelle

Wir haben uns in den bisherigen Überlegungen vornehmlich mit der statistischen Untersuchung von empirisch ermittelten Daten der *Realwelt* (einer *Stichprobe*) beschäftigt und endliche Beobachtungsreihen untersucht. Im letzten Kapitel haben wir auch ein entsprechendes theoretisches *Modell* angeschaut; oft sprechen Statistiker:innen dann von einer *Grundgesamtheit*.

In der Praxis ist es nun so: Die Binomialverteilung beispielsweise ist ein theoretisches Modell, wie die Verteilung einer Grundgesamtheit aussehen könnte. In der Realität werden wir für die empirische Häufigkeitsverteilung bestimmter Zufallsgröße eine Verteilung erhalten, die zwar ungefähr wie eine Binomialverteilung ausschaut, aber eben nur ungefähr – es gibt da und dort kleinere oder größere Abweichungen von der Binomialverteilung. Soweit wir in Abb. 5.4 gesehen haben, können wir zumindest annehmen, dass wir uns an das theoretische Modell umso besser annähern, je größer unsere Stichprobe ist.

In diesem Kapitel fragen wir nun: Inwieweit lassen sich die Ergebnisse einer Stichprobe überhaupt für eine (theoretisch existierende) Grundgesamtheit verallgemeinern und die Ergebnisse aus der Realwelt auf ein Modell übertragen? Wie weit können wir darauf vertrauen, dass die aus den empirischen Daten abgeleiteten Kennwerte auch auf das Modell der Grundgesamtheit zutreffen?

Fragestellungen dieser Art sind Hauptaufgabe der **Induktiven Statistik** (auch: *Schließende* oder *Analytische* Statistik). Zuallererst erinnern wir uns dabei an die wichtigsten Kennwerte einer Stichprobe: Den Mittelwert und die Standardabweichung (siehe Kapitel 3).

Auch für das theoretische Modell, also die Grundgesamtheit, können wir derartige Parameter angeben:

Modell = Grundgesamtheit		**Empirie = Stichprobe**	
Bezeichnung	*Formelzeichen*	*Bezeichnung*	*Formelzeichen*
Erwartungswert	μ	arithmetischer Mittelwert	$\bar{x}$
Standardabweichung	σ	empirische Standardabweichung	s

Tabelle 6.1: Korrespondierende Parameter und Bezeichnungen der (theoretischen) Grundgesamtheit und der (empirischen) Stichprobe

Die Werte in der linken Spalte der Tabelle 6.1 können wir – bis auf wenige Ausnahmefälle – empirisch nicht exakt bestimmen, aber wir können den korrespondierenden Wert der Stichprobe (rechte Spalte) dafür benutzen, die theoretischen Werte zumindest zu *schätzen*. Wenn wir zum Beispiel für μ einen Schätzwert suchen, verwenden wir den empirischen Wert $\bar{x}$, den wir aus einer Stichprobe erhalten haben. Im Sinne der Schätztheorie handelt es sich dabei um einen *Punktschätzer*.

Statistisch gesprochen ist der Erwartungswert μ eine *Zufallsvariable* und der empirische Wert $\bar{x}$ eine *Realisierung* dieser Zufallsvariable.

Wir können uns den «Erwartungswert» auch so vorstellen: Wenn wir eine unendlich große (oder zumindest: ziemlich große) Stichprobe hätten, also zum Beispiel ein Experiment unzählige Male durchgeführt haben, dann sagen wir zum Mittelwert, den wir aus dieser großen Datenmenge erhalten: Erwartungswert.

Aus *einer* Stichprobe erhalten wir zunächst *einen* Schätzwert für den Erwartungswert. Wenn wir *mehrere* Stichproben ziehen, erhalten wir *mehrere* Schätzwerte, sprich: mehrere Realisierungen der Zufallsvariable «Erwartungswert». Wenn wir das oft genug machen (mindestens 30 mal), dann erhalten wir $N \geq 30$ Schätzwerte für den Erwartungswert – und diese Schätzwerte sind normalverteilt[1].

[1] Was wieder mit dem bereits auf Seite 119 erwähnten *Zentralem Grenzwertsatz* zusammenhängt.

Die Weisheit der Vielen
1906 fand im Rahmen der «West of England Fat Stock and Poultry Exhibition», einer regionalen Nutztiermesse in Plymouth, ein Schätzbewerb statt, bei dem das Gewicht eines Ochsen zu raten war. Unter den etwa 800 Personen, die sich daran beteiligten, waren einige Fachleute wie Landwirte und Fleischhauer, aber auch eine große Anzahl von Laien, die einfach ihr Glück im Schätzen versuchen wollten. *Francis Galton* (siehe Fußnote 6, S.82) wertete im Anschluss an den Wettbewerb alle abgegebenen Schätzungen statistisch aus. Eigentlich wollte er zeigen, dass wegen der großen Anzahl von Nicht-Fachleuten das durchschnittliche Ergebnis weit jenseits einer brauchbaren Zahl liegt und man sich daher bei seinen Entscheidungen nicht auf das Urteil von nur durchschnittlich (oder gar unterdurchschnittlich) gebildeten Menschen verlassen sollte. Aber das Gegenteil war der Fall: Das Mittel aus den 787 Schätzwerten betrug mit 542.95 kg beinahe punktgenau das tatsächliche Gewicht von 543.3 kg.

Fundquelle: James Surowiecki. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday.

Schwarmintelligenz anno 1906

6.2 Vertrauensintervalle

Die Abweichung zwischen der Schätzung für einen Parameter und dem wahren Wert dieses Parameters, zum Beispiel die Differenz zwischen Mittelwert der Stichprobe und Erwartungswert der Grundgesamtheit, kann klein oder groß sein. Insbesondere bei einem kleinen Stichprobenumfang wird die Unsicherheit gefühlsmäßig eher größer sein als bei einem großen Stichprobenumfang. Es ist zwar nicht gerade unmöglich aber doch ziemlich unwahrscheinlich, dass wir gleich bei der ersten Stichprobe ins Schwarze treffen.

Um besser auf die Ungenauigkeit der Schätzung eines Parameters des zugrunde liegenden Modells einzugehen, gehen wir so vor:

Wir berechnen aus der Stichprobe nicht nur einen einzigen Wert, z.B. den Mittelwert, sondern wir geben gleich auch ein Intervall rundherum an. Und zwar so, dass das Intervall mit einer Wahrscheinlichkeit von, sagen wir 95%, auch den – uns unbekannten und gar nicht zugänglichen – Erwartungswert der Grundgesamtheit überdeckt. Oder anders ausgedrückt: Wir sind uns zu 95% sicher, dass der Erwartungswert in diesem Intervall enthalten ist.

Zugegebenermaßen ist das so ähnlich wie wenn wir beim Bogenschießen einen

Pfeil auf eine Styroporplatte mit einer «unsichtbaren» Zielscheibe abschießen. Und erst wenn er feststeckt, zeichnen wir die Zielscheibe rundherum. Ausgangspunkt ist der Punkt, wo wir den Pfeil hingeschossen haben. Es ist uns klar, dass wir vermutlich nicht genau ins Zentrum der unsichtbaren Zielscheibe getroffen haben. Daher zeichnen wir einen etwas großzügigeren Kreis, damit wir uns halbwegs sicher sein können, dass das Zentrum der wahren (aber eben unsichtbaren) Zielscheibe auch enthalten ist.

Im Bogensport ist diese Vorgangsweise zwar unüblich, tatsächlich ist es aber eine sehr clevere Methode, aus empirischen Daten auf die Parameter eines unbekannten Modells zu schätzen.

Das gefundene Intervall nennen wir **Konfidenzintervall** (auch: *Vertrauensintervall*)[2]; die Wahrscheinlichkeit, dass wir mit unserer Schätzung richtig liegen, also dass das Intervall, das wir aus der Stichprobe geschätzt haben, den gesuchten Parameter der Grundgesamtheit tatsächlich beinhaltet, ist das **Konfidenzniveau** (auch: *Vertrauenswahrscheinlichkeit*, *statistische Sicherheit* oder *Überdeckungswahrscheinlichkeit* genannt). Das Intervall kann den gesuchten Parameter überdecken oder auch nicht, d.h. wir könnten uns auch irren. Die Wahrscheinlichkeit, dass wir uns irren, wird **Irrtumswahrscheinlichkeit** (auch: *Fehlerwahrscheinlichkeit* oder *Signifikanzniveau*) genannt.

Die Irrtumswahrscheinlichkeit bezeichnen wir üblicherweise mit dem Formelzeichen α und das Konfidenzniveau mit γ, wobei die beiden in Summe immer 1 (bzw. 100%) ergeben müssen[3]: $\gamma = (1 - \alpha)$.

Formal können wir dann ein Konfidenzintervall so ausdrücken:

$$P\,(L \leq \text{Modellparameter} \leq U) = \gamma = (1 - \alpha)$$

mit L . . . *untere* (für «*Lower*») und U . . . *obere Intervallgrenze* («*Upper*»).

Je größer α ist, desto kleiner wird das Konfidenzintervall sein und umgekehrt. Das bringt uns ein bisschen in eine verzwickte Situation: Entweder können wir eine präzise Aussage machen (*Morgen hat es zwischen 10°C und 13°C*), die jedoch höchst unsicher ist, oder eine unscharfe Aussage (*Morgen ist die Temperatur zwischen -10° und +30°*), die sehr zuverlässig eintrifft, aber nicht gerade eine tolle Information enthält.

In der Praxis wählen wir in sozial- und wirtschaftswissenschaftlichen Fragen für α meist 5% oder 10%. (Bei medizinischen oder ingenieurtechnischen Aufgaben

[2] vom lat. *confidere* = vertrauen

[3] Entweder liegen wir richtig, oder nicht. In Summe müssen sich Irrtums- und Vertrauenswahrscheinlichkeit daher auf 1 (bzw. 100%) ergänzen.

wird die Irrtumswahrscheinlichkeit in der Regel kleiner angesetzt, z.B. mit $\alpha = 1\%$ oder 0.1%).

Wenn wir jetzt zum Beispiel aus einer Stichprobe ein Konfidenzintervall für den Erwartungswert μ der zugehörigen Grundgesamtheit schätzen wollen, und wir wählen für $\alpha = 10\%$ bzw. $\gamma = 90\%$, so bedeutet dies: Wir fühlen uns zu 90% «sicher», dass das Intervall, das wir aus der Stichprobe geschätzt haben, den Erwartungswert der Grundgesamtheit enthält. Oder anders ausgedrückt: Wenn wir aus 100 Stichproben jeweils die Konfidenzintervalle bestimmen, ist in 90 derartigen Intervallen damit zu rechnen, dass der Erwartungswert enthalten ist, in 10 Fällen nicht (Leider wissen wir nicht, in welchen 10 Fällen dies eintrifft). $\alpha = 10\%$ ist somit die Angabe eines zehnprozentigen Risikos, dass man bei der Angabe des Konfidenzintervalls eine falsche Aussage tätigt.

Wir können dies auch grafisch veranschaulichen: Nehmen wir an, wir wollen den Erwartungswert μ einer normalverteilten Zufallsgröße bestimmen. Für unser Gedankenbeispiel gehen wir davon aus, dass μ gleich 50 sei (was wir aber in Wirklichkeit nicht wissen – wir wollen μ ja erst bestimmen). Jetzt ziehen wir mehrere Stichproben und bestimmen deren jeweilige Mittelwerte. Jeder dieser Mittelwerte ist dann ein Schätzwert für μ.

In Abb.6.1 sehen wir zehn verschiedene Mittelwerte $\bar{x}_i$, die wir aus den zehn verschiedenen Stichproben erhalten haben. Dazu ist auch jeweils ein Intervall angegeben, in dem unserer Schätzung nach der tatsächliche Wert für μ enthalten sein sollte. Die Intervalle liegen an verschiedenen Stellen (je nachdem, wo der jeweilige Stichprobenmittelwert $\bar{x}_i$ liegt) und sie sind auch unterschiedlich groß (was u.a. mit der Standardabweichung der jeweiligen Stichprobe zusammenhängt). Neun Intervalle liegen nun so, dass μ tatsächlich vom jeweiligen Intervall überdeckt wird. Bei $\bar{x}_5$ hingegen ist μ nicht im Konfidenzintervall enthalten.

Im konkreten Beispiel beträgt also die Wahrscheinlichkeit dafür, ein Intervall wie jenes um $\bar{x}_5$ zu schätzen, $\alpha = 10\%$ (einer von insgesamt zehn Fällen). Die Gegenwahrscheinlichkeit ist $(1 - \alpha) = \gamma = 90\%$. D.h. Die Überdeckungswahrscheinlichkeit beträgt 90%, die Irrtumswahrscheinlichkeit 10%.

Wie können wir jetzt konkrete Grenzen für das Konfidenzintervall angeben? Hier einige Beispiele:

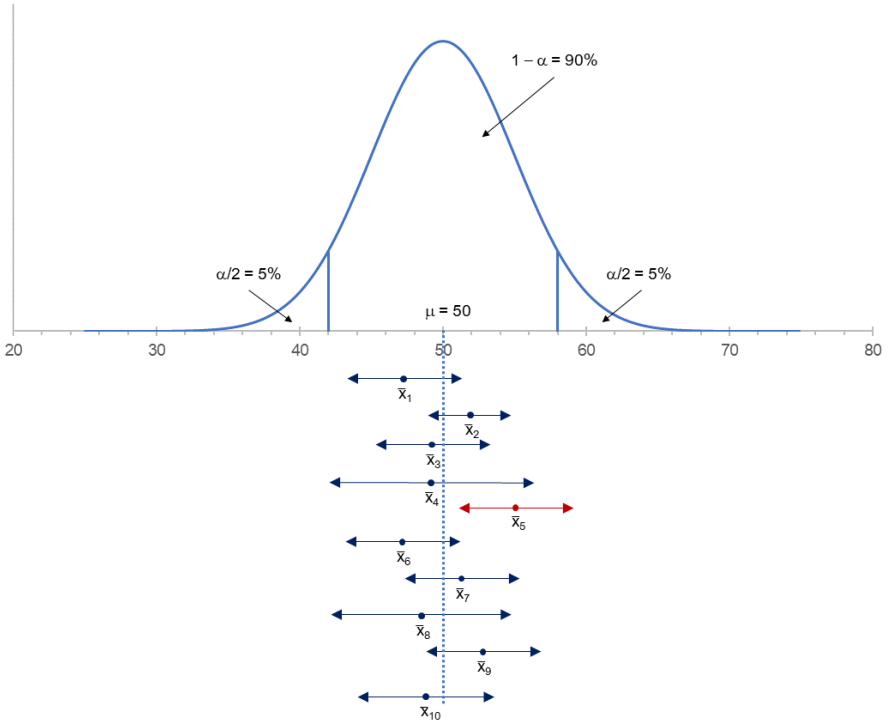

Abb. 6.1: Empirisch erhaltene Konfidenzintervalle einer normalverteilten Zufallsgröße. Deren Erwartungswert beträgt $\mu = 50$ (mit $\sigma = 5$). In einem von 10 Fällen irren wir uns mit dem Konfidenzintervall, aber in 90% liegen wir richtig und μ liegt in dem aus der Stichprobe erhaltenen Intervall. Die Irrtumswahrscheinlichkeit $\alpha = 10\%$ liegt symmetrisch zu jeweils 5% am unteren und oberen Ende der Verteilung.

Schätzung des Konfidenzintervalls für den Erwartungswert μ einer normalverteilten Zufallsgröße

Gegeben sei die Stichprobe einer normalverteilter Zufallsvariablen X. Der Erwartungswert der zugrundeliegenden Grundgesamtheit sei unbekannt und gesucht. Wir bestimmen zunächst einen Schätzwert für den Erwartungswert μ, indem wir uns den arithmetischen Mittelwert $\bar{x}$ ausrechnen. Anschließend berechnen wir auch die empirische Standardabweichung s.

Das Konfidenzintervall für μ ist dann gegeben durch seine untere Grenze L und die obere Grenze U:

$$L = \bar{x} - t_{1-\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}} \tag{6.1}$$

$$U = \bar{x} + t_{1-\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}} \tag{6.2}$$

wobei die Größe $t_{1-\frac{\alpha}{2},n-1}$ das $(1-\frac{\alpha}{2})$-Quantil der so genannten *Studentverteilung* (auch: *t-Verteilung*) ist und aus einem entsprechenden Programm (zum Beispiel *Excel* oder *R*) erhalten werden kann.

Als Eingangsgröße benötigen wir dafür die Anzahl der Elemente in der Stichprobe n sowie die Irrtumswahrscheinlichkeit α. Wählen wir zum Beispiel für $\alpha = 0.05$ und sei $n = 20$, dann beträgt $t = 2.093$.

In *MS Excel* und *LibreOffice Calc* heißt der Befehl zur Berechnung von t `=T.INV.2S(alpha;n-1)`, wobei für `alpha` und `n-1` die jeweiligen Werte für α und n einzusetzen sind.

In *R* lautet der Befehl `qt((1-alpha/2),n-1)`.

Bestimmung der Quantile der T-Verteilung mit *Excel*, *LibreOffice* und *R*.

Der englischer Chemiker und Mathematiker *William Sealey Gosset* (1876-1937) beschäftigte sich um 1900 mit verschiedenen statistischen Analysen und entwickelte die so genannte *Student-* oder *t*-Verteilung.
Die *t*-Verteilung ist – ähnlich der Normalverteilung – über dem Intervall $[-\infty, +\infty]$ definiert, symmetrisch und glockenförmig, hat im Allgemeinen aber eine größere Streuung als die Normalverteilung.
Die *t*-Verteilung ist für uns bei der Bestimmung von Konfidenzintervallen für den Erwartungswert einer Grundgesamtheit wichtig.

Für Fortgeschrittene: Für das Konfidenzintervall benötigen wir mitunter andere Verteilungen als die Normalverteilung.

Das Konfidenzintervall liegt in unserem Fall – wie aus Formeln 6.1 und 6.2 ersichtlich – symmetrisch um den Mittelwert $\bar{x}$ und hat die Länge

$$CI = \frac{2\,t\,s}{\sqrt{n}} \tag{6.3}$$

In *MS Excel* und *LibreOffice Calc* kann mit dem Befehl `=KONFIDENZ.T(alpha;s;n)` (wobei für `alpha` die Irrtumswahrscheinlichkeit, für `s` die Standardabweichung und für `n` die Anzahl der Stichprobenelemente einzugeben sind) die halbe Länge des Konfidenzintervalls ausgerechnet werden, also der Wert $\frac{t\,s}{\sqrt{n}}$. Dieser Wert kann einmal vom Mittelwert abgezogen und einmal dazu addiert werden und wir erhalten nach (6.1) und (6.2) die Grenzen des Konfidenzintervalls.
In *R* können wir den Befehl `t.test(x, conf.level = gamma)` verwenden. Es wird dann gleich mehr berechnet als das Konfidenzintervall (worüber wir uns erst im nächsten Kapitel Gedanken machen), aber unter anderem eben auch das `confidence interval` für die Daten, die im Vektor `x` enthalten sind. Genauer: Je nachdem, was wir für `gamma` eingesetzt haben (z.B. `conf.level = 0.95` oder `conf.level = 0.9`) erhalten wir zum Beispiel das `95 percent confidence interval` oder das `90 percent confidence interval`.
Das 95%-Konfidenzintervall wird dabei als Defaultanwendung gesehen, d.h. wenn wir nur eingeben: `t.test(x)` erhalten wir das 95%-Konfidenzintervall.

Excel-Alternative zur Schätzung des Konfidenzintervalls für den Erwartungswert

Aus Formel 6.3 sehen wir auch: Umso größer der Stichprobenumfang n ist, desto kleiner wird – bei gleichbleibendem α – das Intervall. Kleineres Intervall bedeutet aber: informativere Aussage. Und das bei gleicher Wahrscheinlichkeit.

Es gilt aber auch: Je größer die Standardabweichung s der Stichprobe, desto größer wird das Konfidenzintervall werden.

Für unser Beispiel ($\alpha = 0.05$, $n = 20$, $t = 2.093$) können wir auch leicht rechnen:

$$CI = \frac{2 \cdot 2.093}{\sqrt{20}} \cdot s = 0.936 \cdot s$$

d.h. die Länge des Konfidenzintervalls beträgt 93.6% der Standardabweichung. Verdoppeln wir die Irrtumswahrscheinlichkeit auf $\alpha = 10\%$, dann gilt: $t = 1.729$ und die Länge des Konfidenzintervalls beträgt «nur» mehr 77.3% der Standardabweichung. Wir haben also ein kürzeres Intervall erhalten, aber um den Preis, dass wir uns jetzt in 10 von 100 Fällen irren (vorher war das Intervall größer, aber wir mussten nur in 5 von 100 Fällen mit einem Irrtum rechnen).

Verdoppeln wir hingegen den Stichprobenumfang auf $n = 40$, dann ist t bei

einer Irrtumswahrscheinlichkeit $\alpha = 5\%$ gleich 2.022 und die Länge des Konfidenzintervalls beträgt 64% der Standardabweichung der Stichprobe.

Beispiel 29 *Betrachten wir aus Tabelle 4.1 ausschließlich die Spalte «Größe» und nehmen an, dass diese 20 zufällig ausgewählten Studierenden die übergeordnete Grundgesamtheit «Alle WIBA-Studierenden der FERNFH» repräsentieren.*

Aus den 20 Werten der Stichprobe können wir nach Formel 3.3 einen Mittelwert und nach 3.21 bzw. 3.22 die empirische Standardabweichung ausrechnen. Sie betragen:

$$\bar{x} = 179.5\,cm \quad s = 8.45\,cm$$

Wählen wir für die Irrtumswahrscheinlichkeit $\alpha = 0.05$, dann hat – wie oben angegeben – das entsprechende Quantil der t-Verteilung (für $n = 20$) den Wert $t = 2.093$.

Damit können wir nun die beiden Grenzen des Konfidenzintervalls angeben:

$$L = 179.5 - 2.093 \cdot \frac{8.45}{\sqrt{20}} = 175.5\,cm$$
$$U = 179.5 + 2.093 \cdot \frac{8.45}{\sqrt{20}} = 183.5\,cm$$

D.h. wir nehmen an, dass der tatsächliche Mittelwert aller WIBA-Studierenden irgendwo im Intervall zwischen 175.5 und 183.5 cm liegt.

Mathematisch-statistisch formuliert bedeutet das eben erhalten Ergebnis

$$P\,(175.5\,cm \leq \mu \leq 183.5\,cm) = 95\%$$

was wir «mit Worten» so ausdrücken können:

Erhalten wir für den Parameter μ ein 95%-Konfidenzintervall von $[175.5; 183.5]$, so bedeutet das: *Die Wahrscheinlichkeit, dass der wahre Wert von μ im Intervall $[175.5; 183.5]$ enthalten ist, beträgt 95%*. Oder: *Zögen wir 100 Stichproben und bildeten jeweils das Konfidenzintervall, so würden 95 Intervalle μ enthalten und fünf nicht*.

Diese letzte Aussage lässt uns auch umgekehrt schließen: *Wenn wir aus einer Stichprobe für μ ein 95%-Konfidenzintervall von $[175.5; 183.5]$ erhalten, kann der Erwartungswert der Grundgesamtheit, aus der diese Stichprobe stammt, auch außerhalb dieses Intervalls liegen, also beispielsweise 172 oder 196 cm betragen. Die Wahrscheinlichkeit, dass dies passiert, ist zwar relativ klein (nämlich 5%), aber doch möglich.* Etwas genauer können wir auch sagen: *Im Schnitt wird in 2.5% aller Fälle der wahre Erwartungswert unterhalb von 175.5 cm liegen und in 2.5% aller Fälle oberhalb von 183.5 cm*.

Aufgabe 24 Betrachte aus Tabelle 4.1 ausschließlich die Spalte «Gewicht» und nimm an, dass diese 20 zufällig ausgewählten Personen eine Stichprobe der Grundgesamtheit «Alle WIBA-Studierenden der FERNFH» sind. Gib ein Intervall an, das mit 95%-iger Wahrscheinlichkeit den Erwartungswert für das durchschnittliche Gewicht der WIBA-Studierenden beinhaltet.

Aufgabe 25 Gegeben sind für die Jahre 2005 bis 2014 die Anzahl polizeilich angezeigter Fälle in Österreich:

Jahr	*Anzeigen*	*Jahr*	*Anzeigen*
2005	604 229	2010	534 351
2006	588 229	2011	539 970
2007	592 636	2012	547 764
2008	570 952	2013	546 396
2009	589 961	2014	527 692

In welchen Jahren lag die Anzahl der Anzeigen außerhalb des 95%-Konfidenzintervalls (bezogen auf den Durchschnitt der Jahre 2005-2014)?

Schätzung des Konfidenzintervalls für den Erwartungswert μ einer normalverteilten Zufallsgröße bei Vorliegen einer großen Stichprobe

In den Formeln 6.1 und 6.2 kann man bei großen Stichproben, so ab $n > 30$, die t-Verteilung auch durch eine Normalverteilung ersetzen. Das Konfidenzintervall für μ ist dann gegeben durch die Grenzen

$$L = \bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \tag{6.4}$$

$$U = \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \tag{6.5}$$

Der Koeffizient z ist der entsprechende $(1 - \frac{\alpha}{2})$-Quantilswert der *Normalverteilung*. Als Eingangsgröße benötigen wir nur die Irrtumswahrscheinlichkeit α.

Die Länge des durch (6.4) und (6.5) gegebenen Intervalls ist:

$$CI = \frac{2\,z\,s}{\sqrt{n}} \tag{6.6}$$

Wir können jetzt auch die Intervalllänge CI vorgeben und einen (Mindest-)Wert für den Stichprobenumfang n abschätzen:

$$n > \left(\frac{2\,z}{CI} \cdot \hat{s}\right)^2 \tag{6.7}$$

In *MS Excel* und *LibreOffice Calc* heißt der Befehl zur Berechnung von z `=NORM.S.INV(1-alpha/2)`.

In *R* lautet der Befehl `qnorm(1-alpha/2)`.

In *MS Excel* und *LibreOffice Calc* kann die halbe Länge des Konfidenzintervalls auch direkt ausgerechnet werden. Der Befehl dazu lautet `=KONFIDENZ.NORM(alpha;s;n)`. Diesen Wert muss man einmal von $\bar{x}$ abziehen und einmal dazuzählen und erhält so die obere und untere Grenze des Konfidenzintervalls.

Bei großen Stichproben kann bei der Berechnung des Konfidenzintervalls die Student-Verteilung durch eine Normalverteilung ersetzt werden.

Wie gut wir mit unserer Schätzung liegen, ist jetzt davon abhängig, wie gut wir – vor der Beobachtung der Stichprobe – die Standardabweichung der Stichprobe abschätzen können. (Weil wir die Standardabweichung selbst auch schätzen, bezeichnen wir sie mit $\hat{s}$). Manchmal ist das gar nicht so einfach. Vielleicht ist es ja einfacher abzuschätzen, wie groß der kleinste und größte Wert der Stichprobe ungefähr sein werden. Daraus können wir nach 3.17 die Spannweite Δ ausrechnen (siehe S.69) und daraus wiederum die Standardabweichung abschätzen:

$$\hat{s} = \frac{\Delta}{6} \tag{6.8}$$

(Wie wir auf diese Faustformel kommen? Auf Seite 121 haben wir festgestellt, dass bei einer Normalverteilung ca. 99.7% aller Daten – also schon ziemlich viele – innerhalb eines Bereichs im Abstand von maximal drei Standardabweichungen links und rechts vom Erwartungswert liegen. Macht zusammen sechs Standardabweichungen...).

Beispiel 30 *Wir wollen das 95%-Konfidenzintervall für den Erwartungswert der Körpergröße aller WIBA-Studierenden angeben und dabei soll die Intervalllänge nicht größer als 5 cm sein, d.h. der Erwartungswert soll auf 2.5 cm genau geschätzt werden. Wie groß sollte die Stichprobe mindestens sein, damit wir das schaffen?*

Aus Erfahrung können wir annehmen, dass kein:e Studierende:r kleiner als 150 cm und keine:r größer als 200 cm sein wird. (Und wenn doch, dann sind es nur ganz, ganz wenige). Daraus ergibt sich:

$$\hat{s} = \frac{200 - 150}{6} = \frac{50}{6} = 8.33$$

Für ein Konfidenzniveau von 95% können wir außerdem angeben (z.B. aus Excel): $z =$

1.96, *und somit:*

$$n > \left(\frac{2 \cdot 1.96}{5} \cdot 8.33\right)^2 = 42.7$$

D.h. damit wir ein 95%-Konfidenzintervall für den Erwartungswert der Körpergröße erhalten, das nicht größer als 5 cm ist, sollten wir mindestens 43 Personen in die Stichprobe aufnehmen.

Aufgabe 26 Wir wollen mittels Umfrage herausfinden, mit wie viel Stunden Schlaf unsere Studierenden während des Semesters pro Nacht auskommen (müssen). Konkret wollen wir letztlich durch ein 90%-Konfidenzintervall den Erwartungswert der Schlafdauer auf eine Viertelstunde genau schätzen können. Methodisch bedienen wir uns dabei einer einfachen WhatsApp-Umfrage, d.h. wir ersuchen die Studierenden, uns ein WhatsApp mit den «Schlafstunden» der letzten Nacht zu schicken. Wir erwarten eine maximale Rücklaufquote von 20%, d.h. nur jede:r fünfte Studierende:r, die wir einladen mitzutun, wird uns eine WhatsApp-Nachricht zurückschicken.
Wie viele Personen sollten wir einladen, an der Untersuchung teilzunehmen?

Schätzung des Konfidenzintervalls für den Anteilswert einer Grundgesamtheit

Nicht immer geht es um die direkte Angabe von Mittel- oder Erwartungswerten. Manchmal interessiert uns auch der *Anteil* oder *Prozentsatz* der Merkmalsträger, die eine bestimmte Eigenschaft haben. Wir nennen das dann auch **Anteilswert** bzw. auch: *Quote*). Beispiele sind Ausschussquoten in Produktionsbetrieben, Arbeitslosenquoten, Marktanteile, Einschaltquoten, der Frauenanteil in den Leitungsgremien österreichischer Unternehmen, der Anteil von Personen mit einer bestimmten Meinung oder der Anteil von roten Schokolinsen in einer Packung m&m etc.

Mathematisch ist der Anteilswert definiert als Quotient «Teilgruppengröße durch Gesamtgruppengröße» und lautet für eine Stichprobe mit $n =$ Elementen:

$$\hat{p} = \frac{x}{n} \tag{6.9}$$

mit $x =$ Anzahl der «Treffer», also die Anzahl der Elemente, die die interessierende Eigenschaft haben[4].

[4]Wenn dir bei dieser Formel eine Ähnlichkeit zur Formel 2.5 für die relative Häufigkeit auffällt, ist das kein Zufall…

$\hat{p}$ wird oft in Prozent angegeben. Das ˆ über dem p soll andeuten, dass es sich hier um einen Kennwert aus einer Stichprobe handelt. In den allermeisten Fällen haben wir ja nur eine Stichprobe zur Verfügung. Aus ihm schätzen wir den Anteilswert p der Grundgesamtheit. Es werden zum Beispiel 1 000 Personen (Eltern von Kindern bis 14 Jahren) zur Wiederaufnahme des Schulbetriebs während der Coronakrise befragt und daraus abgeleitet, wie die Mehrheit der österreichischen Eltern darüber denkt. Angenommen, 520 der Befragten – also mehr als die Hälfte – begrüßen die Öffnung der Schulen, wie «sicher» ist die Aussage «Die Mehrheit der Eltern ist für eine Schulöffnung während Corona» bzw. wie hoch ist der Anteilswert in der Grundgesamtheit, wenn wir einen Schätzwert aus der Stichprobe von $\hat{p} = 0.52$ haben? Um diese Frage zu beantworten, können wir wieder ein Konfidenzintervall heranziehen. In dem Fall hat es die beiden Grenzen:

$$L = \hat{p} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \tag{6.10}$$

$$U = \hat{p} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \tag{6.11}$$

Beispiel 31 52 *von* 100 *Befragten sprachen sich in einer Umfrage für eine Öffnung des Schulbetriebs trotz Coronakrise aus. Gib ein* 95%*−Konfidenzintervall für den Anteilswert der zugehörigen Grundgesamtheit an.*

Für $\alpha = 0.05$ *ist* $z = 1.96$ *und mit* $n = 100$ *und* $\hat{p} = \frac{52}{100} = 0.52$ *ergibt sich:*

$$L = 0.52 - 1.96\sqrt{\tfrac{0.52(1-0.52)}{100}} = 0.422$$
$$U = 0.52 + 1.96\sqrt{\tfrac{0.52(1-0.52)}{100}} = 0.618$$

Der Anteilswert der Grundgesamtheit liegt demnach zu 95% *zwischen* 42.2% *und* 61.8%. *D.h. wir können – aus Sicht der Statistik – trotz der* 52% nicht *davon sprechen, dass eine* Mehrheit *der Eltern für eine Schulöffnung ist. Es könnten auch «nur»* 42% *sein.*

Aufgabe 27 Welcher Anteil $\hat{p}$ aus einer Stichprobe mit $n = 183$ Personen müsste sich für einen Gesetzesvorschlag aussprechen, damit wir mit 95%iger Sicherheit darauf schließen können, dass eine (einfache) Mehrheit der Grundgesamtheit für den Beschluss dieses Gesetzes ist?

Im Zusammenhang mit Marktforschung und Meinungsumfragen wird das durch die Grenzen 6.10 und 6.11 definierte Konfidenzintervall auch **Schwankungsbreite** genannt. Konkret ist damit der Wert gemeint, den wir in 6.10 und 6.11

zum empirisch erhaltenen Anteilswert $\hat{p}$ dazuzählen bzw. von ihm abziehen, also:

$$d = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \tag{6.12}$$

Beispiel 32 *Bei der Befragung von 500 potenziellen Kundinnen und Kunden kommt man zum Ergebnis, dass für ein bestimmtes Produkt ein Marktanteil von 30% zu erwarten ist. Wie groß ist die dabei die Schwankungsbreite?*

Für $\alpha = 0.05$ *ist* $z = 1.96$. $n = 500$ *und* $\hat{p} = 0.30$, *daraus ergibt sich:*

$$d = 1.96 \cdot \sqrt{\frac{0.3\,(1-0.3)}{500}} = 0.040$$

Die Schwankungsbreite beträgt also ± 4 *Prozentpunkte.*

Streng genommen handelt es sich bei 6.10 und 6.11 um Annäherungen an die exakten Formeln. Für unsere Zwecke ist dies aber ausreichend genau, zumindest sofern gilt: $n \geq 100$, $n\hat{p} > 5$ und $n\,(1-\hat{p}) > 5$.

Wenn zwar $n\hat{p} > 5$ und $n\,(1-\hat{p}) > 5$ gilt, aber $n < 100$, dann muss man – sofern das Ergebnis auf 1% genau sein soll – noch eine **Kontinuitätskorrektur** anbringen:

$$L = \left(\hat{p} - \frac{1}{2n}\right) - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \tag{6.13}$$

$$U = \left(\hat{p} + \frac{1}{2n}\right) + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \tag{6.14}$$

Für alle, die von verschiedenen (nicht-industriellen) Arten zu fischen eine Vorstellung haben: Durch einen Punktschätzer (vgl. 126) den wahren Wert eines Modellparameters zu erhalten ist vergleichbar mit Speerfischen. Das Konfidenzintervall hingegen gleicht einem Wurfnetz. Und damit werden die meisten von uns mehr Erfolg haben, insbesondere wenn es darum geht, einen ganz bestimmten Fisch zu fangen.
Dabei ist auch klar: Je größer das Netz ist, das wir dabei verwenden, desto plausibler ist es, dass dieser besondere Fisch im Fang auch enthalten ist…

An dieser Stelle noch der Hinweis, dass wir nicht nur für den Erwartungswert oder den Anteilswert einer Verteilung ein Konfidenzintervall ausrechnen können, sondern auch für jeden anderen Parameter, z.B. für eine Standardabweichung, einen Korrelationskoeffizienten, den Anstieg einer Regressionsgeraden etc. In dieser Lehrveranstaltung genügen uns aber Erwartungs- und Anteilswert als Beispiele.

6.3 Kompatibilitätsintervalle und signifikante Unterschiede

Wir können Konfidenzintervalle auch dazu benutzen, zwei Stichproben dahingehend zu vergleichen, ob sie sich bezüglich eines Parameters nur zufallsbedingt unterscheiden oder **signifikant**.

Wenn letzteres zutrifft (signifikanter Unterschied), dann gehen wir davon aus, dass die beiden Stichproben zu verschiedenen Grundgesamtheiten gehören – sich also auch die dahinterliegenden Grundgesamtheiten unterscheiden. Gehören sie hingegen zur selben Grundgesamtheit, dann unterscheiden sich die Parameter nicht sigifikant sondern **zufällig**.

Wir nennen die Konfidenzintervalle dann auch **Kompatibilitätsintervalle**[5] und entscheiden so:

Wenn sich die Kompatibilitätsintervalle der beiden Stichproben nicht oder nur wenig *überschneiden*, dann stammen die Stichproben vermutlich aus verschiedenen Modellen.

Wenn die Überschneidung größer ist, sind die beiden Stichproben miteinander «kompatibel»[6] und deuten auf dasselbe Modell hin.

[5] Das ist nur ein andere Name. Die Berechnung erfolgt gleich wie in den Abschnitten des Kapitels 6.2 angegebenen Fällen.

[6] Daher der Name Kompatibilitätsintervall